

Averaging of Electron Subtomograms and Random Conical Tilt Reconstructions through Likelihood Optimization

Sjors H.W. Scheres,^{1,*} Roberto Melero,¹ Mikel Valle,² and Jose-Maria Carazo¹

¹Biocomputing Unit, Centro Nacional de Biotecnología - CSIC, Darwin 3, Cantoblanco, 28049, Madrid, Spain

²CICBiogune, Parque Tecnológico de Bizkaia, 48160, Derio-Bizkaia, Spain

*Correspondence: scheres@cnb.csic.es

DOI 10.1016/j.str.2009.10.009

SUMMARY

The reference-free averaging of three-dimensional electron microscopy (3D-EM) reconstructions with empty regions in Fourier space represents a pressing problem in electron tomography and single-particle analysis. We present a maximum likelihood algorithm for the simultaneous alignment and classification of subtomograms or random conical tilt (RCT) reconstructions, where the Fourier components in the missing data regions are treated as hidden variables. The behavior of this algorithm was explored using tests on simulated data, while application to experimental data was shown to yield unsupervised class averages for subtomograms of groEL/groES complexes and RCT reconstructions of p53. The latter application served to obtain a reliable *de novo* structure for p53 that may resolve uncertainties about its quaternary structure.

INTRODUCTION

Modern electron microscopes allow the visualization of individual copies of biological macromolecular complexes in a hydrated environment (Stahlberg and Walz, 2008). Because the electron dose needs to be strictly limited to prevent radiation damage, the resulting images are extremely noisy and averaging over multiple copies of identical complexes is often a prerequisite to obtain reliable structural information. However, because of the high noise levels, averaging approaches in three-dimensional electron microscopy (3D-EM) often rely on comparison with a reference structure, and thereby may become sensitive to bias toward this reference. This problem may be particularly aggravated when the data are not measured in its totality, which is the case for subtomograms or RCT reconstructions. Because in both types of experiment the tilt angle in the microscope is physically limited, one cannot collect the data to completeness and the resulting reconstructions contain empty regions in Fourier space. In this contribution, we present an unsupervised averaging approach that can simultaneously align and classify such 3D reconstructions with missing data regions in Fourier space.

In electron tomography, a 3D density map of a unique object (e.g., an entire cell) is obtained by tilting the sample over a controlled series of angles and combining the corresponding

projection images in a single reconstruction (Hoenger and McIntosh, 2009; Leis et al., 2009). In the case of single-axis tilting, the missing region in Fourier space is wedge-shaped, whereas dual-axis tilting reduces this region to a pyramid. The incompleteness of the data in Fourier space results in serious artifacts in the 3D density maps. However, if multiple copies of the same macromolecular complex can be identified and extracted from this reconstruction, one can average over these so-called subtomograms, provided that their relative orientations can be determined and they can be separated into structurally homogeneous classes. Apart from reducing the high levels of noise, averaging also reduces the real-space artifacts due to the incompleteness of the data by “filling in” the missing data regions (Forster and Hegerl, 2007). The latter is true because those data points that are missing in a particle in a given orientation may be observed for another particle in a different orientation, and vice versa. Moreover, by placing the averaged structures back into the cellular reconstruction, one could obtain the “molecular atlas” of the cell, i.e., the spatial distribution of the different functional states of multiple macromolecular complexes throughout the cell. However, because of the high noise levels of individual subtomograms, their incomplete character, and problems related to model bias, the unsupervised alignment and classification of subtomograms currently still represents an important hurdle for what has been called the *visual proteomics* application of cellular tomography (Nickell et al., 2006).

In single-particle analysis, isolated and purified complexes are imaged in a thin layer of vitreous ice, and multiple projections of assumedly identical copies of the complexes are combined in a single 3D reconstruction (Frank, 2006). Because the molecules adopt unknown orientations on the experimental support, their relative projection directions are typically determined a posteriori by comparison of the experimental images with computer-simulated projections of a 3D reference structure. However, the generation of a suitable initial reference map is a highly complicated task, especially for small (i.e., < 400 kDa) or flexible molecules with no or low symmetry (Llorca, 2005). A proven way to generate 3D reconstructions *de novo* is the random conical tilt (RCT) method, where one records image pairs at two different tilt angles in the microscope (Radermacher, 1988; Radermacher et al., 1986). In the general case that the molecules adopt multiple orientations on the support, RCT processing will produce multiple reconstructions in different orientations and, if the molecules are structurally variable, also in distinct conformations. Moreover, the geometrical design of the RCT experiment results in reconstructions with missing regions in Fourier space that are

cone shaped. Thereby, averaging over RCT reconstructions again involves alignment and classification of incomplete reconstructions, and is conceptually very similar to the problem of subtomogram averaging.

Early work on aligning RCT reconstructions includes Penczek et al. (1994). More recently, the potentials of averaging RCT reconstructions have been illustrated by pioneering work on highly flexible spliceosome complexes by Stark and co-workers (Sander et al., 2006), and also Radermacher employed the RCT technique to study highly flexible complexes (Radermacher, 2009). Nevertheless, to date no specific algorithms for the averaging of RCT reconstructions that take the missing cones into account have been described. More contributions are available on subtomogram averaging, and several efforts to obtain molecular atlases of cells have already been reported (e.g., Brandt et al., 2009; Malmstrom et al., 2009). Initial approaches to subtomogram averaging that did not take the missing wedges into account were observed to suffer from the alignment or classification of artifacts (Walz et al., 1997). More recently, the treatment of the missing wedges has played an increasingly important role. The Baumeister group was the first to propose a scoring function that accounts for missing wedges, the constrained cross-correlation coefficient, which limits the comparison of two maps to those regions in Fourier space where both maps contain data. Although initially only used for subtomogram alignment (Beck et al., 2004; Forster et al., 2005; Frangakis et al., 2002), more recently pairwise calculation of this similarity measure for aligned subtomograms was also employed for classification purposes (Forster et al., 2008). Similar approaches to take the missing wedge into account were also used by other groups. Subramaniam and coworkers used a multiplicative mask for the missing wedge in Fourier space and proposed an iterative procedure of efficient subtomogram alignment based on spherical harmonics analysis, and classification based on hierarchical clustering (Bartesaghi et al., 2008; Liu et al., 2008). Winkler et al. classify and align subtomograms based on constrained cross-correlation using either “alignment by classification” (Dube et al., 1993), or iterative schemes involving multivariate statistical analysis, hierarchical clustering, and realignment of the classes (Winkler et al., 2009). Schmid and Booth proposed a simpler scoring function that also restricts the comparison to the mutually observed regions in Fourier space, but is much faster to calculate (Schmid and Booth, 2008).

Common characteristics of existing approaches for subtomogram averaging are that alignment and classification are often performed in separate steps, and that the resulting averages are calculated in Fourier space by dividing the sum of all images by the number of times that each Fourier component was observed. (The latter has been called *weighted averaging*, and we will use the same term throughout this paper.) Both characteristics may behave suboptimally in specific situations. Depending on the type of structural heterogeneity and the overall shape of the particles, the problems of alignment and classification may be strongly intertwined, while weighted averaging in Fourier space may put too strong emphasis on noisy Fourier components that were observed only a few times in cases where the particles adopt a preferred orientation and the missing regions are not filled up.

Our proposal is to tackle alignment and classification simultaneously through a multireference refinement, and to obtain the

parameters of a statistical data model by maximum likelihood estimation. Similar approaches have proven highly effective for the alignment and classification of structurally heterogeneous two-dimensional (2D) projection data sets in single-particle analysis (Scheres et al., 2005, 2007a). A major benefit of the maximum-likelihood approach is the explicit, statistical description of the abundant experimental noise in the data, which allows one to “marginalize,” i.e., calculate probability-weighted assignments, over the so-called hidden variables. These variables are necessary to solve the problem, but were not observed in the experiment. Here, obvious hidden variables are the class and orientational assignments of every particle. In addition, to explicitly reflect the unobserved character of the missing data in our statistical model, we also treat all Fourier components in the missing data regions as hidden variables. We will derive the corresponding algorithm, where the missing data regions are complemented with the current estimates for the model parameters, and we will show that this algorithm may be employed in a completely unsupervised manner for the alignment and classification of subtomograms or RCT reconstructions.

Results

The Data Model

We model our data set to consist of N 3D reconstructions (particles) X_i , with 3D Fourier transforms X_i , which are described according to the following model in Fourier space:

$$X_i = R_{\Phi_i} A_{\kappa_i} + G_i \text{ for } i = 1, 2, \dots, N \quad (1)$$

where:

- $X_i \in \mathbb{R}^J$ is a J -dimensional vector of complex values $(X_i)_j$ or X_{ij} in short, which can be divided into a vector X_i^o of observed components X_{ij}^o and a vector X_i^m of missing components X_{ij}^m . Each data vector X_i may have a unique pattern of missing components, but the pattern is known for every X_i . We formalize the notion of a missing data generation mechanism by defining a J -dimensional missing data indicator vector w_i , such that $w_{ij} = 1$ for observed elements, and $w_{ij} = 0$ for missing elements.

- A_{κ_i} is one of K unknown 3D structures in Fourier space $A_1, A_2, \dots, A_K \in \mathbb{R}^J$. These are the objects we wish to estimate from the data. κ_i is an unknown, random integer with possible values 1, 2, ..., K , indicating which of the unknown structures corresponds to particle X_i .

- R_{Φ_i} is a transformation matrix, which maps the unknown structure A_{κ_i} on particle X_i . The actual transformation Φ_i for particle X_i is unknown. We parameterize this transformation by ϕ_i , a 6D vector (corresponding to three Euler angles q_{rot} , q_{tilt} , q_{psi} , and three real-space translation coordinates, q_x , q_y , q_z).

- $G_i \in \mathbb{R}^J$ is a J -dimensional vector of unknown, independent Gaussian noise with mean zero and unknown standard deviation σ .

In this model, the data space comprises the entire J -dimensional vector X_i , including the part that remained unobserved, which contrasts with alternative approaches where the data space is restricted to only the observed part of X_i , (e.g., Forster and Hegerl, 2007). We note here that also in the latter case one may devise a maximum-likelihood algorithm, and investigations into alternative ways to deal with preferred orientations within such a data model are currently underway (unpublished results). Also, the assumption of white Gaussian noise in this model may

only be an approximate description of experimental noise in 3D-EM. Still, similar assumptions in maximum-likelihood (ML) classification approaches for single-particle data have been shown to be highly effective (Scheres et al., 2005, 2007a). Moreover, the mathematical framework presented here could easily be adapted to describe colored noise by introducing a resolution-dependent estimate for the standard deviation of the noise. Again, for single-particle analysis such an approach has previously been shown to be effective (Scheres et al., 2007b), but we chose not to implement a similar approach here because of its increased computational complexity.

The Optimization Problem

The task at hand is to estimate A_1, A_2, \dots, A_K from the partially observed data vectors X_1, X_2, \dots, X_N . We view this estimation problem as an incomplete data problem, where the missing data associated with X_i are their unknown transformation Φ_i and class assignment κ_i , as well as the vector of missing Fourier components X_{ij}^m . Note that if the data would have been complete, i.e., if Φ_i , κ_i and X_{ij}^m would have been observed, this problem would be extremely easy to solve by straightforward averaging. Because this is not the case, we deal with the incompleteness of the data by expectation maximization (Dempster et al., 1977). This method provides a general solution for finding maximum likelihood estimates of parameters in probabilistic models that depend on unobserved, or so-called “hidden variables,” which are added to the observed data forming a complete-data vector. The algorithm forms a likelihood function on the complete-data vector, and then overcomes the fact that part of the data is missing by iteratively working with the conditional expectation of the complete-data log-likelihood given the observed data, which is effected using the current fit for the missing parameters.

In this case, the complete data set corresponds to

$$(X_i^o, X_i^m, \Phi_i, \kappa_i) \text{ for } i = 1, 2, \dots, N, \quad (2)$$

and we aim to find those parameters that maximize the joint probability of observing the entire set of observed data vectors $X_1^o, X_2^o, \dots, X_N^o$:

$$L(\Theta) = \sum_{i=1}^N \log \sum_{k=1}^K \int_{\phi} \int_{M_i} P(X_i^o | k, \phi, X_i^m, \Theta) P(k, \phi, X_i^m | \Theta) d\phi dM_i \quad (3)$$

where $\int_{M_i} dM_i$ is a shorthand notation for the integrals from $-\infty$ to ∞ for every missing Fourier component X_{ij}^m in X_i : $\int_{M_i} dM_i = \int_{X_{i1}^m = -\infty}^{\infty} \int_{X_{i2}^m = -\infty}^{\infty} \dots dX_{i1}^m dX_{i2}^m \dots$.

According to the data model in (1), the probability of observing X_i^o for a given combination of the hidden and model parameters is expressed as:

$$P(X_i^o | k, \phi, X_i^m, \Theta) = \left(\frac{1}{2\pi\sigma^2} \right)^J \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^J \left(w_{ij} \| (R_{\phi} A_k)_j - X_{ij}^o \|^2 + (1 - w_{ij}) \| (R_{\phi} A_k)_j - X_{ij}^m \|^2 \right) \right\} \quad (4)$$

We assume that the prior probabilities of all k , ϕ and X_{ij}^m are independent, and that all rotations and all values of the missing Fourier components are equally likely. In addition, we assume that the translational components of ϕ are distributed according to a 3D Gaussian, centered at the origin and with standard deviation ξ , and the proportion of particles belonging to the class is described by values α_k (with $\alpha_k \geq 0$, and $\sum_{k=1}^K \alpha_k = 1$). Thereby, $P(k, \phi, X_{ij}^m | \Theta)$ is expressed as:

$$P(k, \phi, X_{ij}^m | \Theta) = \frac{C\alpha_k}{\xi^3} \exp \left\{ \frac{q_x^2 + q_y^2 + q_z^2}{-2\xi^2} \right\}, \quad (5)$$

where C is a constant to ensure that $\sum_{k=1}^K \int_{\phi} \int_{M_i} P(k, \phi, X_{ij}^m | \Theta) d\phi dM_i = 1$.

From (5) and (6) we see that model parameter set Θ comprises estimates for A_1, A_2, \dots, A_K , $\alpha_1, \alpha_2, \dots, \alpha_K$, σ , and ξ , while as stated above, the hidden variables consist of k , ϕ for every particle and all X_{ij}^m .

The Optimization Algorithm

In the Supplemental Data (available online), we derive the following expectation-maximization algorithm for the optimization of the log-likelihood function in (4). The estimates for the underlying 3D structures may be updated by:

$$A_{kj}^{\text{new}} = \frac{\sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \left\{ (R_{\phi}^{-1} w_{ij}) (R_{\phi}^{-1} X_i^o)_j + (1 - (R_{\phi}^{-1} w_{ij})) A_{kj}^{\text{old}} \right\} d\phi}{\sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) d\phi}, \quad (6)$$

where R_{ϕ}^{-1} denotes the inverse transformation of R_{ϕ} , and $P(k, \phi | X_i^o, \Theta)$ is the conditional probability of k and ϕ in terms of the current model estimates and calculated only over the observed Fourier components X_i^o .

Similarly, the estimate for the standard deviation in the noise may be updated using:

$$(\sigma^{\text{new}})^2 = \frac{1}{NJ} \sum_{i=1}^N \sum_{j=1}^J \sum_{k=1}^K \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \times \left\{ w_{ij} \| (R_{\phi} A_k^{\text{old}})_j - X_{ij}^o \|^2 + (1 - w_{ij}) (\sigma^{\text{old}})^2 \right\} d\phi \quad (7)$$

The updates of α_k and ξ are analogous to their updates in the 2D case (Scheres et al., 2005), and are given by:

$$\alpha_k^{\text{new}} = \sum_{i=1}^N \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) d\phi \quad (8)$$

$$(\xi^{\text{new}})^2 = \frac{1}{3N} \sum_{i=1}^N \sum_{k=1}^K \int_{\phi} P(k, \phi | X_i^o, \Theta^{\text{old}}) \left\{ q_x^2 + q_y^2 + q_z^2 \right\} d\phi \quad (9)$$

Comparison with the ML Algorithm for Single Particles

It is interesting to compare the estimation problem and the optimization algorithm at hand with those for the simultaneous alignment and classification of 2D single-particle projections (Scheres et al., 2005). Although the latter concerned a 2D case

in real space, Θ comprised the same set of model parameters. The main differences between the two cases lie in the hidden variables. In the 2D case, ϕ comprised translational coordinates in only two directions and only a single in-plane rotation, whereas for the case at hand 3D rotations and three translations are to be evaluated. Moreover, the data vectors in the 2D case were completely observed, so that X_i^m did not form part of the hidden variables.

Consequently, also the two optimization algorithms are similar. As for the 2D case, the algorithm derived above calculates model parameter updates as probability-weighted averages over all possible assignments of the hidden variables ϕ and k . However, rather than calculating these probabilities over the entire data vectors, they are calculated only over the observed elements. In this way, the underlying similarity measure of the experimental particles and the average is similar to the constrained cross-correlation coefficient as proposed previously (Forster et al., 2005; Frangakis et al., 2002). Due to the incompleteness of the data vectors, additional terms arise in (7) and (8), compared to their counterparts in the 2D case. To be more concise, for those regions in Fourier space that were unobserved, the data vectors are complemented with the corresponding values of the model at hand (A_{kj}^{old} or σ^{old}). That is, one “fills in” the missing wedges of the incomplete data vectors with the current model estimates in order to generate a pseudo-complete data set. In this way, the iterative algorithm functions in a way that is very similar to the complete-data case and that does not involve any division by the number of times a Fourier component has been observed, as would be the case in approaches based on weighted averaging.

A Regularized Optimization Target

We also implemented an algorithm to optimize a regularized log-likelihood function. Rather than maximizing (4), this algorithm optimizes the following target, which contains an additional empirical term that restraints the differences between the distinct class averages by minimizing the square L2-norms of all respective difference maps:

$$\Lambda(\Theta) = L(\Theta) - \frac{\vartheta}{2\sigma^2} \sum_{k=1}^K \sum_{l=1}^K v_{kl} \sum_{j=1}^J \|A_{kj} - A_{lj}\|^2. \quad (10)$$

Here, θ and v_{kl} are free parameters that determine to what extent the similarity among the different class averages is imposed. We note that this regularization is very similar to the neighbor restraints in the kernel-density self-organizing map (Pascual-Montano et al., 2002). Actually, although our current implementation only employs $v_{kl} = 1$, we note that for a special choice of v_{kl} a neural network-like algorithm may be obtained that is very similar to that of the self-organizing map, but that complements the missing wedges with the model estimates from the previous iteration. Because the regularization term in a way contradicts the goal of clustering (to obtain as different as possible class averages), we typically only employ this term during the first few iterations and gradually decrease the value of θ to zero.

The corresponding algorithm is very similar to the one outlined above (see [Supplemental Data](#) available online). The update formulae for the model estimates A_{kj} and σ contain addition terms compared with the updates given in (7) and (8), while the

updates for α_k and ξ are not affected by the regularization term and remain as given in (9) and (10).

RESULTS

Model Calculations

To test the potentials of the presented ML approach, we first used simulated data to compare their performance with a standard multireference refinement based on constrained cross-correlation coefficients and weighted averaging (CC). The simulated data were generated in a perhaps simplistic manner to represent the idealized case, where the data are in accordance with the underlying statistical model of our approach. Atomic models of GroEL₁₄ and GroEL₁₄GroES₇ complexes (Protein Data Bank [PDB] IDs 1GR5 and 2C7C, respectively) were converted to 2.5 nm resolution electron density maps using the `convert_pdb2vol` program from the Xmipp package (Sorzano et al., 2004). All simulated data sets contained 200 subtomograms, for which GroEL₁₄ or GroEL₁₄GroES₇ were randomly selected with equal probability. These maps were subsequently rotated and translated, various amounts of independent Gaussian noise were added, and a missing wedge was applied in the Fourier domain corresponding to a tilt series from -60° to 60° . Random translations were sampled from a uniform distribution between -3 and $+3$ voxels in all three directions. Random rotations were sampled from an approximately even distribution of three Euler angles with random perturbations. To simulate different degrees of preferred orientations, Euler triplets where the second Euler angle was between 60° and 120° or between 40° and 140° were omitted from the angular distributions. This corresponded to a moderate and a more severe preferred orientation of the particles with an increasing lack of information along the seven-fold axis of the GroEL complexes. The direction of these preferred orientations corresponds to the worst-case scenario, i.e. it coincides with the direction in which the differences between the two structures are most different.

Unsupervised alignment and classification runs with these model data sets comprised refinements with $K = 2$ references and an angular sampling of 15° . All refinements were started from the weighted averages of two random subsets of the data in random orientations, and all runs were stopped after 40 iterations. Preliminary experiments without the additional regularization term on the log-likelihood function (11) showed suboptimal classification behavior for both the ML and the CC algorithms, which appeared to get stuck in local minima at the initial iterations of the optimization process. Therefore, in all subsequent ML and CC runs we employed a regularization scheme where the regularization parameter θ was initially set to $5N/K^2$, and then linearly decreased to 0 in the first 5 iterations. Then, multiple runs of the ML and the CC approaches were compared for data sets with the moderate preferred orientation and signal-to-noise ratios (SNRs) ranging from 0.002 to 0.1 (Figure 1). Except for the lowest tested SNR, where both approaches failed, the ML approach gave purer classes than the CC approach over the whole SNR range. Notably, at SNRs higher than 0.007 the ML approach gave almost completely pure classes for all runs performed, whereas even for SNRs as high as 0.1 the CC approach still failed in some of the runs. In the latter cases, typically all images were assigned to a single

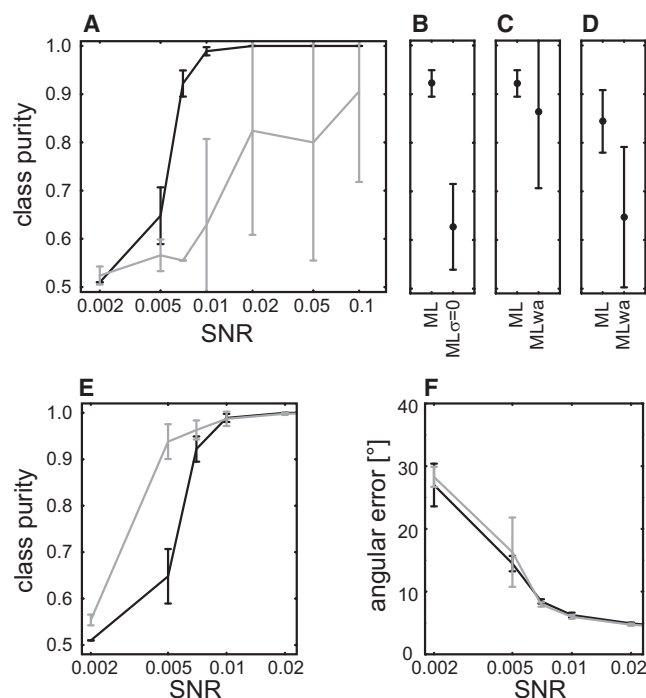


Figure 1. Model Calculations

(A) Class purities (see below) for the ML (black) and CC (gray) approach to unsupervised multireference refinement of simulated data sets with varying amounts of noise. Averages and standard deviations (error bars) are calculated over five runs that were started from different random class and orientation assignments.

(B) As in A, for simulated data with SNR = 0.007, comparing the ML approach with a modified ML approach (MLσ = 0) where the standard deviation of the noise was fixed at a value 100 times lower than the actual noise level.

(C) As in B, comparing the ML approach with a modified ML approach (MLwa) where the marginalization over the missing data regions was replaced by a weighted-averaging approach.

(D) As in C, but for data with a more severe preferred orientation of the particles.

(E) Class purities for the ML multireference refinements (as in A, black) and for the ML refinements with perfectly aligned particles (gray).

(F) Errors in the angular assignments of the groEL₁₄ complexes in the ML multireference refinements (black) and the ML refinements with structurally homogeneous subsets (gray). To calculate class purities and angular errors, class and angular assignments were based on the maximum of the probability distribution of each particle. Then, as given on p. 549 of (Tan et al., 2006), class purity was defined as

$$\frac{1}{N} \sum_{k=1}^2 \max_t m_{kt},$$

with m_{kt} being the number of particles from true structure t assigned to class k . Angular errors were calculated as the average angular distance between the three vectors of the coordinate systems defined by the perfect Euler angles and the angles obtained in the ML refinements. All 14 equivalent orientations due to the D7 symmetry of the groEL₁₄ complex were taken into account. Note that, as a consequence of this high symmetry the angular errors at low SNR (almost 30°) correspond to near-random angular assignments.

class and one class was left empty. Such behavior was not observed for the ML approach, but we note that although these solutions result in low class purities, they can be easily discerned from correct solutions.

To gain a better understanding of the ML algorithm, we then performed two additional experiments for the data set with a SNR of 0.007, which appeared to be on the limit of the amount of noise the ML approach can handle for this phantom. The first experiment was designed to test the effect of marginalizing over all orientations and classes, i.e., the calculation of probability-weighted class and orientation assignments. This was done by performing ML runs with σ fixed at a value 100 times smaller than the actual noise in the data. This effectively converts the probability distributions into approximate delta functions, yielding an algorithm that is closely related to the CC approach (cf. Sigworth, 1998), although it still complements the missing data with the current model estimates in each M-step. These runs gave significantly worse classification results than the true ML approach (Figure 1B), and detailed analyses of the convergence paths of both runs (not shown) indicated that the runs with near-delta function probability distributions have an increased tendency to get stuck into local minima at early stages of the optimization process.

The second experiment was designed to test the effect of marginalizing over the missing data regions. This was done by adapting the ML algorithm to use weighted averaging rather than marginalization over the missing data regions. That is, rather than complementing the missing wedges with the model estimates from the previous iteration, the adapted ML algorithm divides every Fourier component in the average by the number of times it was observed, while it still calculates probability-weighted averages over all orientation and class assignments. For data with the moderate preferred orientation, complementing the missing wedges yielded somewhat better class purities than weighted averaging, but the differences were smaller compared with the run with near-delta functions (Figure 1C). Much more pronounced differences were observed for data with the more severe preferred orientation (Figure 1D). For the latter data, the missing wedges upon alignment of all particles do not fill up Fourier space in the average, while for the moderate preferred orientation at least several experimental observations were available for all Fourier components. Detailed analyses of these runs revealed that, in particular for the data with the severe preferred orientation, the numerical instabilities caused by division by relatively small numbers negatively affected the optimization paths of the weighted-average algorithm. In several cases, this algorithm was even observed to diverge from reasonable solutions at intermediate stages to much worse solutions at the end of the runs.

Finally, we performed two experiments to evaluate the alignment and classification properties of our algorithm separately. First, we provided the perfect alignment parameters of the simulated data sets and applied the ML algorithm to classify these data while keeping the angular assignments fixed. As expected for the less complicated task of classifying perfectly aligned particles, purer classes could be obtained at relatively low SNRs compared with the multireference refinement schemes described above (Figure 1E). Second, we applied the ML algorithm (with $K = 1$) to align the structurally homogeneous subsets consisting of only the GroEL₁₄ complexes of the data sets described above. Remarkably, despite the less complicated task of aligning structurally homogeneous data, over the entire SNR range the resulting errors in the angular assignments

were very similar to those observed in the multireference refinements (Figure 1F). Even for SNRs where the multireference refinements gave relatively impure classes (e.g., SNR = 0.005), the errors in the angular assignments were comparable to those obtained for structurally homogenous data. Our interpretation of these results is that, in this particular case, the alignment may suffer relatively little from remaining heterogeneity in the data because of the structural resemblance of the GroEL₁₄ and GroEL₁₄GroES₇ complexes.

GroEL-GroES: Subtomogram Averaging

Unsupervised alignment and classification of experimental subtomograms was tested using a previously published test data set on purified groEL and groEL/groES complexes (Forster et al., 2008) (see [Experimental Procedures](#) for details). Several particles were discarded upon visual inspection due to remaining density for gold particles. The remaining 780 subtomograms were aligned using a single reference, which was obtained in an unsupervised manner by calculating the (weighted) average of all subtomograms in random orientations. After 25 iterations, the ML approach converged to a barrel-like structure with a clear seven-fold symmetry component (not shown). Subsequently, a second run with $K = 3$ references was started from the averages of three random subsets of the data, maintaining the angles obtained in the first run. In this run an angular sampling of 10 was employed, while the angular search range was limited to angular distances within 50 from the angles of the first run. As for the simulated data, the empirical regularization term θ was initially set to $5N/K^2$, and then linearly decreased to 0 in 5 iterations.

Central slices through the resulting maps after 25 iterations are shown in Figure 2A. Of the particles assigned to class 1, 90% came from tilt series from preparations of groEL; likewise, 86% and 100% of the particles assigned to classes 2 and 3, respectively, came from tilt series from preparations of GroELGroES complexes. The average map of class 1 showed an open, barrel-like structure with clear seven-fold symmetry and was interpreted as GroEL. Also the map of class 2 showed a barrel-like structure with seven-fold symmetry, but with partially closed, additional density on one of its ends, and this map was interpreted as GroELGroES. The map of class 3 showed a smaller structure with less clear indications for seven-fold symmetry. This class could not be directly related to GroEL or GroELGroES complexes, and possibly coincides with the classes that were tentatively interpreted as groES aggregates by (Forster et al., 2008). Because the maps of classes 1 and 2 both showed clear indications of seven-fold symmetry, C7 symmetry was enforced, and subsequent fitting of the atomic structures of these complexes (PDB IDs 1GR5 and 2C7C) in UCSF Chimera (Pettersen et al., 2004) resulted in excellent fits with cross-correlation coefficients of 0.88 and 0.81, respectively (Figures 2B and 2C).

P53-DNA: RCT Reconstruction Averaging

The usefulness of the ML approach for the unsupervised averaging of RCT reconstructions was illustrated by its application in an ongoing project of our labs on the structural characterization of human p53. Two strikingly different quaternary structures have previously been proposed for murine (Okorokov et al., 2006) and human (Tidow et al., 2007) p53 complexes. In both cases, these structures were obtained by refinement of initial

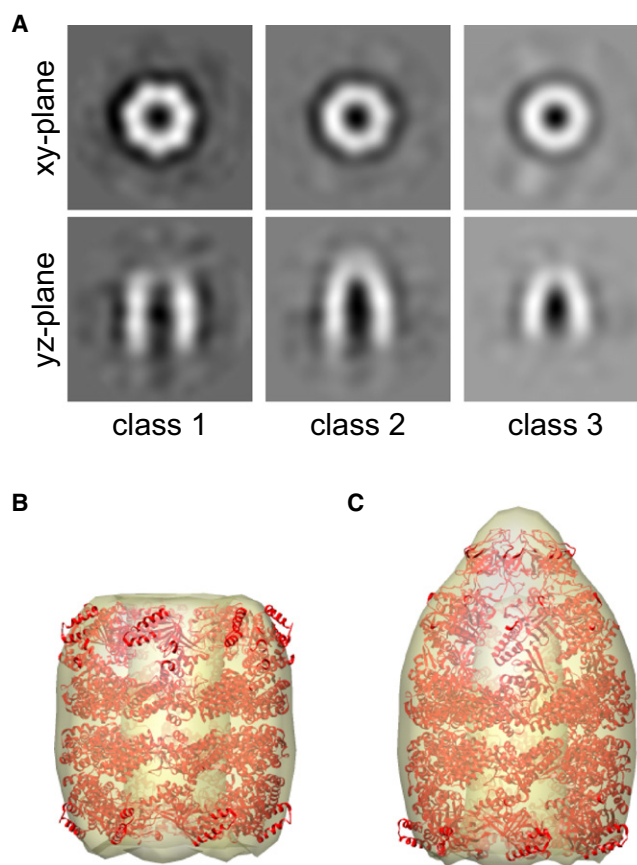


Figure 2. GroEL/GroES Subtomogram Averaging

(A) Central slices through XY and YZ of the three class averages obtained for the data set of subtomograms of groEL and groEL/groES complexes.

(B) Symmetrized average subtomogram for class 2 with fitted atomic model of groEL₁₄.

(C) Symmetrized average subtomogram for class 3 with fitted atomic model of groEL₁₄groES₇.

models that were based on angular reconstitution (common-lines reconstruction) (Van Heel, 1987). In this work, we set out to generate an alternative de novo model for p53 using the RCT technique, and subsequently refined this model using the EMAN package (Ludtke et al., 1999).

A data set of 2521 tilt pair particles of a stabilized p53 mutant lacking its 33 C-terminal residues (see [Experimental Procedures](#)) in complex with a dsDNA probe was processed using standard procedures in the Xmipp package (Scheres et al., 2008) to produce 40 RCT reconstructions in a range of different orientations. We then employed the ML algorithm outlined above to align these reconstructions in a completely unsupervised manner. Given the limited number of available 3D maps, the averaging of these reconstructions was performed using only a single reference (i.e., $K = 1$). Starting from a 4 nm low-pass filtered, weighted average of all RCT reconstructions in random orientations, 13 iterations of the ML algorithm were performed with an angular sampling of 15. The resulting average showed indications for two-fold symmetry (cf. Figures 3A and 3D), which was subsequently enforced on the averaged map. The symmetrized

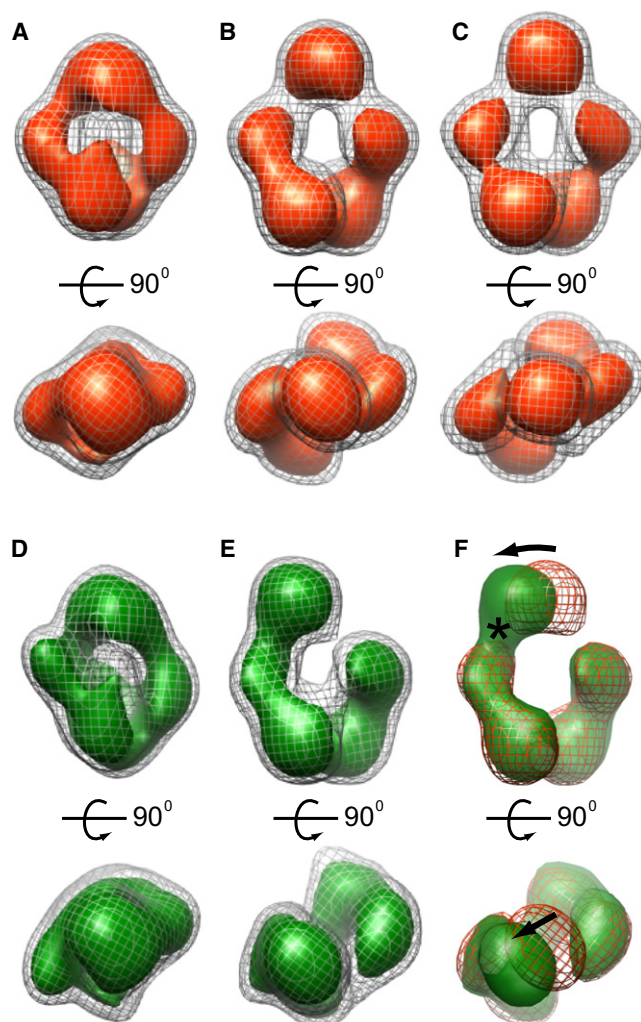


Figure 3. p53-DNA RCT Reconstruction Averaging

(A) RCT average with imposed C2 symmetry.
 (B) Refinement of the RCT average against the untilted data set, imposing C2 symmetry.
 (C) Refinement of a common-lines generated model against the untilted data set, imposing C2 symmetry.
 (D) RCT average without imposing any symmetry.
 (E) Refinement of the RCT average against the untilted data set, without imposing any symmetry.
 (F) Comparison of the refined RCT models with (orange mesh) or without (green) imposing C2 symmetry shows a departure of the top domain from the symmetry axis (arrows) and a gain in linker density with one of the core dimers (asterisk) in the asymmetric model.
 All C2-symmetrized maps are shown in orange; all C1 maps are shown in green. Maps are thresholded at 50% of the expected mass in green and orange, and at 100% of the expected mass in (mesh) gray. The side views (top rows) are related to the top views (bottom rows) by a 90-degree rotation, and the putative two-fold symmetry axis is perpendicular to the plane of the top views.

map (Figure 3A) displayed a large mass at the top of the complex and four masses at the bottom, with weaker linkers connecting the top and bottom parts. This organization has remarkable similarities with the previously proposed quaternary structure for human p53, where the top part was interpreted as the tetrameri-

zation domain and the bottom part as dimers of DNA-binding core domain dimers (Tidow et al., 2007).

Subsequent refinement of the average RCT reconstruction with C2 symmetry in EMAN against a second data set of 7600 (untilted) particles further increased the notion of a dimer of dimers (Figure 3B). Two additional observations validate the overall organization of this structure. First, a very similar C2-structure (Figure 3C) was obtained in a completely independent manner by refinement of a model based on common-lines reconstruction (startAny option in EMAN). Second, refinement of the asymmetric RCT model without imposing any symmetry yielded a structure that was virtually identical to the C2-structure at the bottom part, while only the mass at the top part moved slightly away from the two-fold axis (Figures 3E and 3F).

DISCUSSION

We have presented a new approach for averaging 3D maps with missing regions in Fourier space. The novelty of our approach lies in three general aspects. First, whereas many existing approaches separate alignment and classification, we tackle both problems simultaneously through multireference refinements. Second, we use maximum-likelihood estimation to calculate our model parameters as weighted averages over all possible orientation and class assignments, rather than assigning a single optimal orientation and class based on a noisy (constrained) cross-correlation function. And third, we explicitly treat the notion of incomplete experimental data in our statistical model by treating the Fourier components in the missing data regions as hidden variables. The latter results in an algorithm where one fills in the missing Fourier components with the corresponding components of the current model estimates, rather than dividing each Fourier component by the times it has been observed (as in previously proposed approaches based on weighted averaging). This gives rise to a “conservative” algorithm, which changes the model much more slowly compared with weighted averaging for those regions in Fourier space that are sparsely observed among all particles. The user may tune this degree of conservatism, because one could perform multiple iterations of marginalization over the missing data regions for every iteration of marginalization over the orientations and classes. In the limit of an infinite number of iterations of marginalization over the missing data regions, this algorithm then becomes equivalent to its weighted averaging counterpart. Although not used in the calculations presented in this paper, this option was implemented in the program and comes at virtually no additional computational cost.

The concept of complementing the missing data with the available model estimates is not restricted to the ML approach presented here. Also in approaches based on constrained cross-correlation, weighted averaging may be replaced by complementing the missing Fourier components with the current model estimates. Such changes would probably require only minor adjustments to the code of existing programs and may provide more robust algorithms for cases where the particles adopt preferred orientations. In this light, we also note the work of Bostina et al., who fill up the missing data regions of their aligned subtomograms with the corresponding values of the weighted average of all data, prior to application of standard

3D classification tools that lacked the notion of missing data (Bostina et al., 2007).

Initial tests of the presented algorithm were performed with simulated data that were in agreement with the assumptions of the underlying data model. For data with white, Gaussian noise in a wide range of SNRs, the proposed ML algorithm was shown to outperform a multireference refinement approach based on weighted averaging and constrained cross-correlation, which is in good agreement with similar observations for single-particle refinements (Scheres et al., 2005; Sigworth, 1998).

For both the ML and the CC approaches, the classification results were greatly improved using an empirical regularization term. We note that, in general, regularization may be employed to prevent overfitting or to solve ill-posed problems. Often, regularization terms tend to smooth the energy landscape of a target function, thereby lowering the number of local minima. Also in this case, our understanding of this effect is that the regularization prevents the multireference refinement runs from getting stuck in local minima at early stages of the refinement. During the initial iterations, the quality of the references is still relatively low due to suboptimal alignment parameters, and separation of the images into distinct groups at these stages may converge to mediocre solutions. The regularization term that acts during these initial iterations enforces similarity among the references, so that these local minima may be overcome. Starting from initially random assignments of rotations and classes, the ML algorithm was then shown to be capable of yielding simultaneous classification and alignment in a completely unsupervised manner.

As explained above, the main differences of the ML algorithm with conventional approaches lie in the probability-weighted averaging over all possible orientational and class assignments on one hand and the filling of the missing regions with the current model estimates on the other hand. In experiments designed to separate these two types of marginalization, in particular the probability-weighted averaging over the orientation and class assignments appeared to have a large effect on the refinement results. Complementing the missing data regions seems to be of secondary importance, although this effect became more noticeable for data with severe preferred orientations where the missing regions were not filled up in the averages.

The experimental data set of GroEL/groES subtomograms that we used was previously published by Foerster et al., who used constrained cross-correlation to align all particles against a reference based on a low-pass filtered crystal structure, and then classified these data using hierarchical clustering based on the same similarity measure (Foerster et al., 2008). We obtained similar classification results, but in a completely unsupervised manner. Moreover, the maps for the groEL₁₄ and groES₁₄groES₇ complexes obtained in our work appear to agree somewhat better with the fitted crystal structures than those presented by Foerster et al., although a quantitative comparison is not available at this point. Apart from the theoretical considerations outlined above, such an improvement could also be explained by the simultaneous alignment and classification in our refinement scheme, whereas Foerster et al. separated alignment against a single reference from subsequent classification and did not re-align the resulting classes.

The reference-free alignment of p53-DNA RCT reconstructions showed that, despite a low number of noisy RCT reconstructions, the ML approach was capable of generating an averaged map with remarkable similarities to a previously observed superdomain organization of human p53 (Tidow et al., 2007). Apart from illustrating the potentials of the ML approach in generating de novo models for subsequent single-particle refinement, these results also have relevance for p53 structural biology. Previously, two strikingly different quaternary structures were proposed based on common-lines EM-reconstructions. Tidow et al. (2007) used negative-stain EM to study a stabilized mutant of human full-length p53. Enforcing C2 symmetry, they obtained models with similar domain organizations as those shown in Figure 3, either for free p53 tetramers and for complexes with dsDNA. (Okorokov et al., 2006) used cryo-EM on full-length murine p53, and enforced D2 symmetry to obtain a model with the shape of a hollow, skewed cube. As p53 may adopt an intrinsically flexible quaternary organization to fulfill its multiple functionalities in the cell, it is possible that both EM reconstructions represent relevant oligomerization states (Okorokov and Orlova, 2009). On the other hand, some caution may be at place, because the imposition of an incorrect symmetry in the generation and refinement of common-lines models may lead to important artifacts, and the inherent flexibility combined with its small size (200 kDa) further complicate (cryo-) EM analysis of p53 samples. Our RCT averaging experiments yielded the first de novo p53 model that is not based on common lines. Moreover, whereas the previously proposed common-lines models could not be obtained without imposing symmetry, our model could be generated and refined in the absence of any symmetry.

Summarizing, this contribution provides a general framework for maximum-likelihood refinement of 3D maps with missing data, and the usefulness of this approach has been illustrated using both simulated and experimental data of subtomograms and RCT reconstructions. As has happened previously with similar approaches the field of single-particle analysis, it is likely that the underlying statistical data model may need to be adapted to describe a wide variety of experimental characteristics. Apart from the theoretical considerations about the data model outlined above, neighboring particles from the crowded environment of the cell could interfere with the correct normalization of subtomograms, as was previously observed for single-particle analyses (Scheres et al., 2009). In addition, densities in cellular subtomograms for neighboring complexes, gold particles, or membranes that are not described by the data model may lead to artifacts in the alignment and classification process, whereas real-space masks (to remove such artifacts) may lead to additional dependencies in Fourier space and to underestimation of the noise variance. To facilitate further testing and possible adaptation in the field, the presented algorithm has been made freely available through its implementation in the open-source Xmipp package (Sorzano et al., 2004). Extrapolating our experience with maximum likelihood refinement of structurally heterogeneous single-particle projection data (Scheres et al., 2005, 2007a), we expect that this implementation may provide a useful tool for unsupervised alignment and classification of electron subtomograms or RCT reconstructions.

EXPERIMENTAL PROCEDURES

GroEL/groES Subtomogram Data

The tomographic data on groEL/groES mixtures used in this work was published previously (Forster et al., 2008). These authors collected single-axis tilt series on purified protein samples of GroEL and GroELGroES complexes mixed with a 10 nm BSA-colloidal gold suspension. Carbon-coated grids were vitrified by plunge freezing, and low-dose images were collected using a Tecnai G2 Polara microscope (FEI, Eindhoven, The Netherlands) equipped with a 2k × 2k CCD camera. The tilt series ranged from −65 to 65 with a 2–2.5 increment, defocus values between 4 and 7 μm were used, and the pixel size at specimen level was 0.6 nm. Manual particle selection yielded several thousand subtomograms, which were downsampled to 32 × 32 × 32 voxels with a final voxel size of 1.2 nm. From this data set, 793 subtomograms were selected based on a cross-correlation criterion with the average of all subtomograms aligned against a low-pass-filtered crystal structure of the GroELGroES complex (Xu et al., 1997).

P53 Electron Microscopy

A deletion mutant of human p53, lacking its 33 C-terminal residues (NTC, p53₁₋₃₆₀) and with four stabilizing mutations in the core domain (M133L/V203A/N239Y/N268D) (Nikolova et al., 1998), was expressed in *E. coli*, and purified as described elsewhere (Veprintsev et al., 2006). Complexes with a double-stranded DNA probe containing the TGFA response element were prepared at a high concentration (10 μM) to avoid free p53. These samples were simultaneously separated on molecular weight and cross-linked with glutaraldehyde using the Grafix method (Kastner et al., 2008). For all fractions, the interaction of the DNA probe with p53 was confirmed by electrophoretic mobility shift assay, and the size of the complex was confirmed using SDS-PAGE. Only fractions containing p53 tetramers in complex with DNA were selected for EM. Negative staining was performed with uranyl formate in carbon sandwich technique over Quantifoil grids. Micrographs were taken in tilted pairs, at tilt angles of 0° and 45°, under low-dose conditions on a JEOL JEM-2200FS electron microscope. Images were recorded on a 4k × 4k CCD camera at a magnification of 64,305 ×, corresponding to a 2.1 Å pixel size. A total of 2521 pairs of particles were manually selected, downsampled to a pixel size of 8.4 Å, and processed using standardized RCT protocols in Xmipp (Scheres et al., 2008) to generate 40 reconstructions of 30 × 30 × 30 voxels. Similarly, a second data set consisting of 7600 untilted particles was obtained using the same protein sample, but incubated with a dsDNA probe containing the GADD45 response element.

Implementation

The ML algorithm outlined above was implemented in the *xmipp_ml_tomo* program of the Xmipp package (Sorzano et al., 2004). The same program also implements a standard multireference refinement scheme based on weighted averaging and maximum constrained cross-correlations (CC). In addition, although we did not derive the corresponding formulas here, the program implements a modified maximum-likelihood algorithm that employs weighted averaging to calculate the reference structures. The latter two implementations depend on an additional free parameter in the form of a threshold value to prevent divisions by zero, which was set to 0.1 in all calculations. Although the program reads 3D maps in single-file SPIDER format only, the Xmipp package provides utilities to convert to and from other common formats like MRC (Crowther et al., 1996) or EM (Hegerl, 1996). Missing data regions may be defined as cones, wedges, or pyramids, and the program may also be used to align 3D maps without missing regions. The 6D integrations over all orientational and translational assignments are performed exhaustively, although the orientational integrations may optionally be limited to a given angular distance from a set of initial orientations for all particles. The integrations over all orientations are replaced by discrete Riemann sums over approximately even distributions of the three Euler angles. It is noteworthy that these distributions are oriented differently according to a small random perturbation at every iteration to prevent the optimization from getting stuck in local minima. The program has been parallelized with near-linear efficiency for the numbers of processors employed in this study, using a hybrid scheme of distributed and shared memory parallelization by the combined use of the message passing interface (MPI) and threads. The computational cost of the

implementation scales linearly with the number of particles, the number of classes and the number of sampled rotations. In addition, it scales approximately linearly with the number of voxels in the 3D maps. The runs with the model data took approximately 7–8 hr wall clock time each, using 20 3GHz Intel Xeon processors in parallel. The groEL runs took in total 17 hr on 128 2.3GHz IBM Power PC processors, and the alignment of the p53 maps took 20 min on 20 3GHz Intel Xeon processors.

SUPPLEMENTAL DATA

Supplemental Data include a Supplemental Approach and can be found with this article online at [http://www.cell.com/structure/supplemental/S0969-2126\(09\)00411-0](http://www.cell.com/structure/supplemental/S0969-2126(09)00411-0).

ACKNOWLEDGMENTS

We thank Michael Stoelken, and Stephan Nickell and Friedrich Foerster for useful comments, the latter for providing the groEL/groES data, Alan Fersht for providing p53 samples, and David Gil and Melissa Lazaro for help with p53 data collection. We are grateful to the Barcelona Supercomputing Center (BSC-CNS) for providing computer resources. Funding was provided by the Spanish Ministry of Science (CSD2006-00023, BIO2007-67150-C03-1/3) and the Comunidad de Madrid (S-GEN-0166-2006), the European Union (FP6-502828), the U.S. National Heart, Lung and Blood Institute (R01 HL070472). The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung and Blood Institute.

Received: August 7, 2009

Revised: October 2, 2009

Accepted: October 22, 2009

Published: December 8, 2009

REFERENCES

- Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., and Subramaniam, S. (2008). Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* 162, 436–450.
- Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W., and Medalia, O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* 306, 1387–1390.
- Bostina, M., Bubeck, D., Schwartz, C., Nicastro, D., Filman, D.J., and Hogle, J.M. (2007). Single particle cryoelectron tomography characterization of the structure and structural variability of poliovirus-receptor-membrane complex at 30 Å resolution. *J. Struct. Biol.* 160, 200–210.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* 136, 261–271.
- Crowther, R.A., Henderson, R., and Smith, J.M. (1996). MRC image processing programs. *J. Struct. Biol.* 116, 9–16.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum-likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.* 39, 1–38.
- Dube, P., Tavares, P., Lurz, R., and van Heel, M. (1993). The portal protein of bacteriophage SP1: a DNA pump with 13-fold symmetry. *EMBO J.* 12, 1303–1309.
- Forster, F., and Hegerl, R. (2007). Structure determination in situ by averaging of tomograms. *Methods Cell Biol.* 79, 741–767.
- Forster, F., Medalia, O., Zauberman, N., Baumeister, W., and Fass, D. (2005). Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA* 102, 4729–4734.
- Forster, F., Pruggnaller, S., Seybert, A., and Frangakis, A.S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* 161, 276–286.
- Frangakis, A.S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R., and Baumeister, W. (2002). Identification of macromolecular

- complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. USA* 99, 14153–14158.
- Frank, J. (2006). *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (New York: Oxford University Press).
- Hegerl, R. (1996). The EM Program Package: A Platform for Image Processing in Biological Electron Microscopy. *J. Struct. Biol.* 116, 30–34.
- Hoenger, A., and McIntosh, J.R. (2009). Probing the macromolecular organization of cells by electron tomography. *Curr. Opin. Cell Biol.* 21, 89–96.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* 5, 53–55.
- Leis, A., Rockel, B., Andrees, L., and Baumeister, W. (2009). Visualizing cells at the nanoscale. *Trends Biochem. Sci.* 34, 60–70.
- Liu, J., Bartesaghi, A., Borgnia, M.J., Sapiro, G., and Subramaniam, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. *Nature* 455, 109–113.
- Llorca, O. (2005). Introduction to 3D reconstruction of macromolecules using single particle electron microscopy. *Acta Pharmacol. Sin.* 26, 1153–1164.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128, 82–97.
- Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 460, 762–765.
- Nickell, S., Kofler, C., Leis, A.P., and Baumeister, W. (2006). A visual approach to proteomics. *Nat. Rev. Mol. Cell Biol.* 7, 225–230.
- Nikolova, P.V., Henckel, J., Lane, D.P., and Fersht, A.R. (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl. Acad. Sci. USA* 95, 14675–14680.
- Okorokov, A.L., and Orlova, E.V. (2009). Structural biology of the p53 tumour suppressor. *Curr. Opin. Struct. Biol.* 19, 197–202.
- Okorokov, A.L., Sherman, M.B., Plisson, C., Grinkevich, V., Sigmundsson, K., Selivanova, G., Milner, J., and Orlova, E.V. (2006). The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J.* 25, 5191–5200.
- Pascual-Montano, A., Taylor, K.A., Winkler, H., Pascual-Marqui, R.D., and Carazo, J.M. (2002). Quantitative self-organizing maps for clustering electron tomograms. *J. Struct. Biol.* 138, 114–122.
- Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53, 251–270.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Radermacher, M. (1988). Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microsc. Tech.* 9, 359–394.
- Radermacher, M. (2009). Chapter 1 Visualizing functional flexibility by three-dimensional electron microscopy reconstructing complex I of the mitochondrial respiratory chain. *Methods Enzymol.* 456, 3–27.
- Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1986). A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. *J. Microsc.* 141, RP1–RP2.
- Sander, B., Golas, M.M., Makarov, E.M., Brahm, H., Kastner, B., Luhrmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* 24, 267–278.
- Scheres, S.H., Valle, M., Nunez, R., Sorzano, C.O., Marabini, R., Herman, G.T., and Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* 348, 139–149.
- Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J., and Carazo, J.M. (2007a). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods* 4, 27–29.
- Scheres, S.H., Nunez-Ramirez, R., Gomez-Llorente, Y., San Martin, C., Eggermont, P.P., and Carazo, J.M. (2007b). Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure* 15, 1167–1177.
- Scheres, S.H., Nunez-Ramirez, R., Sorzano, C.O., Carazo, J.M., and Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* 3, 977–990.
- Scheres, S.H., Valle, M., Grob, P., Nogales, E., and Carazo, J.M. (2009). Maximum likelihood refinement of electron microscopy data with normalization errors. *J. Struct. Biol.* 166, 234–240.
- Schmid, M.F., and Booth, C.R. (2008). Methods for aligning and for averaging 3D volumes with missing data. *J. Struct. Biol.* 161, 243–248.
- Sigworth, F.J. (1998). A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* 122, 328–339.
- Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., and Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* 148, 194–204.
- Stahlberg, H., and Walz, T. (2008). Molecular electron microscopy: state of the art and current challenges. *ACS Chem. Biol.* 3, 268–281.
- Tan, P.N., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining* (Addison-Wesley).
- Tidow, H., Melero, R., Mylonas, E., Freund, S.M., Grossmann, J.G., Carazo, J.M., Svergun, D.I., Valle, M., and Fersht, A.R. (2007). Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. USA* 104, 12324–12329.
- Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21, 111–123.
- Vepintsev, D.B., Freund, S.M., Andreeva, A., Rutledge, S.E., Tidow, H., Canadillas, J.M., Blair, C.M., and Fersht, A.R. (2006). Core domain interactions in full-length p53 in solution. *Proc. Natl. Acad. Sci. USA* 103, 2115–2119.
- Walz, J., Typke, D., Nitsch, M., Koster, A.J., Hegerl, R., and Baumeister, W. (1997). Electron tomography of single ice-embedded macromolecules: three-dimensional alignment and classification. *J. Struct. Biol.* 120, 387–395.
- Winkler, H., Zhu, P., Liu, J., Ye, F., Roux, K.H., and Taylor, K.A. (2009). Tomographic subvolume alignment and subvolume classification applied to myosin V and SIV envelope spikes. *J. Struct. Biol.* 165, 64–77.
- Xu, Z., Horwich, A.L., and Sigler, P.B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 388, 741–750.